Enrichment of a LINE subfamily in a single chromosomal region in *Peromyscus*

David H. Kass,1 John A. Peppers,2 Mary Maltbie,2 Robert J. Baker2

1Department of Biology, Eastern Michigan University, Ypsilanti, Michigan 48197, USA
2Department of Biological Sciences, Texas Tech University, Lubbock, Texas 79409, USA

Received: 15 December 1997 / Accepted: 20 February 1998

LI/NES are a class of transposable elements greater than 5 kb in length consisting of $10^5$-$10^6$ copies, most of which are heterogeneously 5’ truncated (Voliva et al. 1983). These elements are not randomly inserted in the genome, and generally correspond to the A-T rich chromosomal G bands, as well as accumulate on the sex chromosomes, as observed in the human (Korenberg and Pykowski 1988), European house mouse (Boyle et al. 1990), and deer mouse (Wichman et al. 1992; Baker and Kass 1994).

Sequence analysis of L1 elements suggests few original source or master genes generated the numerous L1 elements (Deininger et al. 1992). As mutations accumulate in the master gene, the more recently diverged species would share these mutations and yield concerted variations and hence subfamilies. LINE subfamilies have been described by restriction endonuclease cleavage patterns (Jubier-Maurin et al. 1985; Kass et al. 1992), sequences of 5’ Ntots (Padgett et al. 1988), diagnostic nucleotides (Jurka 1989), and stretches of sequence variability within an open reading frame (ORF; Pascale et al. 1990; Kass et al. 1992). Two highly divergent LINE (L1) subfamilies (L1Pm62 and L1Pm55) have been identified that coexist in the genomes of *Peromyscus*, exhibiting differing restriction fragment length variants (RFLVs) on Southern blots with DNA isolated from various *Peromyscus* species (Kass et al. 1992). Sequences of the elements (subcloned L1-hybridizing fragments from a *P. maniculatus* genomic library) representing these subfamilies demonstrate 29% divergence from each other within ORF-2, which is unusually high for intraspecific LINEs compared with mice and rats (Martin et al. 1985; Soares et al. 1985). Comparisons of second and third position nucleotides by codon alignment between L1Pm62 and L1Pm55 support the view that these subfamilies were derived from different source genes (Kass et al. 1992). A comparative sequence analysis of several *Peromyscus* LINEs depicts L1Pm62 and L1Pm55 as relatively "old", pre-dating the divergence of two recent LINE-1 lineages (Casavant et al. 1996). The lack of a conserved ORF additionally supports the ancient origin of the L1-62 and L1-55 subfamilies (Kass et al. 1992) and is indicative of retropositional inactivity during the course of speciation within this genus. However, the maintenance of conserved restriction sites that follow phylogenetic trends (Kass et al. 1992) suggests these subfamilies have more recent evolutionary histories.

L1Pm55 appeared to be the oldest sequence of elements isolated from *P. maniculatus* and branches as an offshoot from the other *Peromyscus* L1 sequences from a derived molecular phylogeny (Casavant et al. 1996). L1Pm55 was estimated at 100 copies per haploid genome by a genomic Southern blot (Kass et al. 1992), representing less than 1% of total LINE repeats. The same blot demonstrated 2 less intense EcoRI fragment in species of the *P. leucopus* (0.7 kb) and *P. truei* (1.7 kb) species groups, indicative of fewer copies or high sequence divergence. The use of alternative restriction enzyme for Southern blots also demonstrated concerted variations, primarily at the species group level, always with a more intense band in species of the *P. maniculatus* species group. Surprisingly, this suggests a more recent origin of this apparently retropositionally inactive subfamily. We therefore elected to analyze the chromosomal organization of L1Pm55 to gain insight into the evolutionary history of this enigmatic subfamily.

Fluorescence in situ hybridization (FISH) as previously described (Wichman et al. 1992; Baker and Kass 1994) was used to analyze chromosomes of *P. leucopus, P. maniculatus*, and *P. melanotis*. The L1 hybridizing probe (pDK55) is a 1.5 kb fragment subcloned from a *P. maniculatus* genomic library (Kass et al. 1992). Although two small regions within pDK55 (80 bp and a 27 bp GC-rich region) are not observed in L1Pm62 or L1 sequences from other mammals (Kass et al. 1992), it is highly improbable that the observed FISH patterns are due to homologies to these short DNA stretches. A BLAST search of the database (Altschul et al. 1990) identified no corresponding sequences to the 27 bp region. Various sequences located on different mouse and human chromosomes were identified that corresponded to the 80 bp region, and any significant sequence identities (80%) that were observed included only various 20 bp stretches of this region. One microgram of the clone (plasmid plus insert) was biotinylated by nick translation, and the pDK55 probe was hybridized in the presence of sheared *E. coli* DNA. Hybridization under the conditions used (Baker and Kass 1994) should detect sequences greater than 70% in similarity. The hybridized probes were detected with fluorescent-conjugated avidin (Vector Labs), followed by two amplifications consisting of alternate treatments of biotinylated goat anti-avidin (Vector Labs). Digital enhancement of images and pseudo-G banding were accomplished by using software from OtCOR Images. Repeated hybridizations for each species confirmed the results.

The L1-55 subfamily is primarily located on a single pair of homologous chromosomes (Fig. 1) within *P. maniculatus* and *P. melanotis*, both of the *P. maniculatus* species group, and in *P. leucopus*. The distinguishable FISH pattern in *P. leucopus* further supports the subfamily classification of L1-55, as opposed to one or a few aberrant element(s) strictly within the *P. maniculatus* genome. The accompanying pseudo-G band spreads support the existence of this subfamily on the same chromosome in these species. The use of *P. melanotis* further denotes Chr 12 as the residence for L1-55, based on G banding patterns (Committee for the Standardization of Chromosomes of *Peromyscus*, 1994). L1-55 had not been observed by FISH in *P. eremicus* (data not shown), a member of a separate subgenus (Haplomyomys) of *Peromyscus*, either as a consequence of the level of divergence, low copy number, or its absence.

Several mechanisms for the non-random chromosomal ar-
rangement of transposable elements have been proposed (Wichman et al. 1992). These mechanisms, though, account for the accumulation of LINEs on G bands and the sex chromosomes. However, alternative evolutionary mechanisms need to be explored to explain the predominant localization of the L1-55 subfamily on a single chromosome (autosomal) region.

The lack of an open reading frame in L1Pm55 and the observed chromosomal pattern suggest that the concerted variation of the L1-55 family in Peromyscus species was not the result of either a recently evolved master gene or recent activation of a suppressed source gene. An example of a replication-competent, but relatively inactive, LINE has been suggested to exist in the murine genome (Pascale et al. 1993). Although LINE insertions are generally non-random (Korenberg and Rykowski 1988), it would be difficult to conceive that this LINE subfamily would specifically integrate primarily (or possibly only) in this one chromosomal region.

Concerted evolution of interspersed repetitive sequences was originally suggested to occur by unequal crossing over, gene conversion, and DNA transposition (Dover 1982). Although retroposition has been considered the primary force of L1 evolution (Deininger et al. 1992) unequal crossing over has been proposed to explain the narrow genomic positioning of repetitive sequences (Smith 1976). The narrow positioning of this LINE subfamily suggests a role of unequal crossing over in the evolution of LINEs. Alternatively, a LINE may integrate into another repeated sequence, such as a long complex repeat unit (LCRU), as observed in Mus (Nasir et al. 1991), and subsequently be amplified by unequal crossing over. The alternative mechanism, gene conversion, may also play a role. Gene conversions have been observed in SINEs (Kass et al. 1995; Batzer et al. 1995). However, in these cases, it was proposed that the conversion was driven by a cDNA derived from a source/master gene. In the L1-55 example, gene conversions would more likely occur owing to proximity.

L1-55 represents the only known mammalian LINE subfamily predominantly localized to a single chromosomal region. This supports the original presumptions of Dover that unequal crossing over and gene conversion are mechanisms involved in the evolution of interspersed repeat sequences, although they apparently remain secondary to retroposition.

Acknowledgments. D.H. Kass was supported by a Provost's New Faculty Award and a University Spring/Summer Research Award from Eastern Michigan University. J.A. Peppers was supported in part by an Achievement Reward for College Scientists Scholarship.

References

Casavan NC, Sherman AN, Wishman MA (1996) Two persistent LINE-1 lineages in Peromyscus have unequal rates of evolution. Genetics 142, 1289-1298
Nasir J, Maceonchoe MK, Brown SD (1991) Co-amplification of L1 line elements with localized low copy repeats in Giemsa dark bands:
Implications for genome organization. Nucleic Acids Res 19, 3255–3260
Padgett RW, Hutchison CA III, Edgell MH (1988) The E-type 5' motif of mouse L1 elements: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. Nucleic Acids Res 16, 739–749
Pascale E, Valle E, Furano AV (1990) Amplification of an ancestral mammalian L1 family of long interspersed repeated DNA occurred just before the primate radiation. Proc Natl Acad Sci USA 87, 9481–9485